1982

Nonparametric Analysis of Right Censored Data with Multiple Comparisons

Hwei-Weng Shih
Utah State University

NONPARAMETRIC ANALYSIS OF RIGHT CENSORED DATA WITH MULTIPLE COMPARISONS

by

Hwei-Weng Shih

A report submitted in partial fulfillment
of the requirements for the degree
of
MASTER OF SCIENCE
in
Applied Statistics
(Plan B)

UTAH STATE UNIVERSITY
Logan, Utah
1982

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to my major professor, David L. Turner, for his patient guidance, and help in organizing the paper. Thanks are also extended to my committee members, Dr. Ronald V. Canfield and Dr. Gregory Jones, for their help and criticism.

I would also like to express my sincere thanks to my parents for their encouragement and support in my graduate studies at Utah State University.

Hwei-Weng Shih

TABLE OF CONTENTS

LIST OF TABLES

ABSTRACT

Nonparametric Analysis of Right Censored

Data with Multiple Comparisons

by

Hwei-Weng Shih, Master of Science

Utah State University, 1982

Major Professor: Dr. David L. Turner

Department: Applied Statistics

This report demonstrates the use of a computer program written in FORTRAN for the Burroughs B6800 computer at Utah State University to perform Breslow's (1970) generalization of the Kruskal-Wallis test for right censored data. A pairwise multiple comparison procedure using Bonferroni's inequality is also introduced and demonstrated. Comparisons are also made with a parametric F test and the original Kruskal-Wallis test. Application of these techniques to two data sets indicate that there is little difference among the procedures with the F test being slightly more liberal (too many differences) and the Kruskal-Wallis test corrected for ties being slightly more conservative than Breslow's test statistic.

(30 pages)

CHAPTER I

INTRODUCTION

Statistical relationships between variables must sometimes be estimated from incomplete data or data which has been censored. Censoring occurs when an experiment is stopped before the event of interest occurs. When this happens the recorded data do not provide direct information about the event. In this paper we shall consider only samples censored on the right. This means that the only information about the censored observations is their total number and the fact that each is greater than some known value. For example, if we were studying survival time of a patient or animal under a set of experimental conditions, the data would be analyzed while some patients or animals are still alive. According to Lagakos (1979) the analysis of censored data can be used to obtain as much information as an uncensored experiment would yield.

A fundamental problem in many life testing problems is a comparison of the survival-time distributions from two or more samples of censored data. Norman Breslow (1970) reviews a generalization of Wilcoxon's statistic for comparing two populations as proposed by Gehan (1965) for use when the observations are subject to arbitrary right censorship. Breslow also discusses Mantel's (1967) further generalization to the case of arbitrarily restricted observations, or left and right censorship. Both Mantel and Gehan base their calculations on the permutation distribution of the statistic, conditional on the observed censoring pattern for the combined sample.

Breslow (1970) extended Gehan's generalization of Wilcoxon's test to allow for testing the equality of K continuous distribution functions when observations are subject to arbitrary right censorship. Breslow's generalization is an extension of the Kruskal-Wallis test, and is the "state of the art" nonparametric test of equality of K groups with possibly differing distributions for the censoring variables.

Breslow's development of this extended Kruskal-Wallis test involves some very complicated formulae. He gives two "easy" approximations but even these would be very laborious to compute.

This report demonstrates the use of a computer program written for the Burroughs B6800 computer which translates Breslow's formulae into a form which may actually be used. A pairwise multiple comparison procedure using Bonferroni's inequality is also developed and demonstrated. Two sets of data will be analyzed using Breslow's procedure and the Bonferroni multiple comparison procedure. Comparisons will also be made with the parametric (F test) procedure for the case of data from exponential distributions. The nonparametric Kruskal-Wallis test for uncensored data will also be applied using the modifications for tied data discussed in Ott (1977).

CHAPTER II

METHODOLOGY

In this report the major method used to analyze the hypothesis is Breslow's generalization of the Kruskal-Wallis test. In addition to Breslow's method, the Kruskal-Wallis test is also used to test the hypothesis that $K \geq 2$ populations are identical using modifications when there are ties in the data. An F test for the case of two exponential distributions is also performed. Comparisons will then be made among the various methods when results are known using a set of generated or Monte Carlo Data. The methods are then applied to a real set of data.

A Special Comparison for Two Exponential Distributions

Let the two exponential distributions with parameters λ_1 and λ_2 have probability density function

$$f(X_{ij};\ \lambda_i) = \frac{1}{\lambda_i} \exp(-X_{ij}/\lambda_i) \qquad i = 1,\ 2;\ j = 1,\ 2, \ldots,\ n_i$$

Then

$$\hat{\lambda}_1 = \frac{\sum_{j=1}^{n_i} X_{ij}}{\sum_{j=1}^{n_i} {}_{ij}} = \frac{X_{i\cdot}}{\delta_{i\cdot}}$$

is the maximum likelihood estimate of λ_i, $i = 1, 2$, where X_{ij} equals the true value or censored value depending on whether δ_{ij} equals 1 (uncensored) or 0 (censored). Then

$$R = \frac{\hat{\lambda}_1}{\hat{\lambda}_2}$$

is an F distributed random variable with $2\delta_{1\cdot}$ and $2\delta_{2\cdot}$ degrees of freedom. This result may be used to test H_0: $\lambda_1 = \lambda_2$.

The Wilcoxon Two Sample Rank Sum Test and the Kruskal-Wallis Test

The Wilcoxon rank sum test provides a nonparametric test of the hypothesis that two populations are identical, since the experimenter has obtained two samples from possibly different populations, and we wish to use a statistical test to see if we can reject the null hypothesis that the two populations are identical. That is, we wish to detect differences between the two populations on the basis of random samples from those populations. An approach to the two-sample problem is to rank the combined data from lowest to highest. We let R_1 denote the sum of the ranks for sample 1. R_1 can take on values ranging from $n_1(n_1 + 1)/2$ to $(n_1 + n_2)(n_1 + n_2 + 1)/2 - n_2(n_2 + 1)/2$. Intuitively, if R_1 is close to either extreme, we would have evidence to reject the null hypothesis that the two populations are identical, since sample 1 would then be all close to the bottom or the top of the ranked distribution.

The concept of a rank sum test was extended to a comparison of more than two populations by Kruskal and Wallis (1952). The $K \geq 2$ random samples have been obtained from each of K possibly different populations, and we want to test the null hypothesis that all of the populations are identical against the alternative that some of the populations tend to furnish greater observed values than other populations.

To perform the test, the $K \geq 2$ samples are combined into a single ordered sample, then ranks are assigned to the sample values from the smallest value to the largest, without regard to which population each value came from. Let N denote the total number of observations,

$$N = \sum_{i=1}^{K} n_i$$

where n_i is the number of observations from sample i. Let $R(X_{ij})$ denote the rank assigned to X_{ij}, R_i be the sum of the ranks assigned to the ith sample,

$$R_i = \sum_{j=1}^{n_i} R(X_{ij}) \qquad i=1, 2, \ldots, K.$$

Note that

$$\sum_{i=1}^{K} R_i = 1 + 2 + \ldots + N = \frac{N(N + 1)}{2} .$$

If there are several observations tied or equal to each other, the average of their ranks is assigned to each of the tied observations.

The large sample approximation for the test statistic T is based on the fact that R_i is the sum of n_i random variables. So the mean and variance of R_i are given by

$$E(R_i) = \frac{n_i(N + 1)}{2}$$

and

$$Var(R_i) = \frac{n_i(N + 1)(N - n_i)}{2}$$

Therefore

$$\frac{R_i - E(R_i)}{\sqrt{Var(R_i)}}$$

is approximately distributed as a standardized normal random variable when n_i is large enough. Thus

$$\left[\frac{R_i - E(R_i)}{\sqrt{Var(R_i)}}\right]^2 = \frac{\{R_i - [n_i(N + 1)/2]\}^2}{n_i(N + 1)(N - n_i)/12}$$

is approximately distributed as a chi-square random variable with one degree of freedom. If the R_i were independent of each other the distribution of the sum

$$T = \sum_{i=1}^{K} \frac{\{R_i - [n_i(N + 1)/2]\}^2}{n_i(N + 1)(N - n_i)/12}$$

could be approximated using the chi-square distribution with K degrees of freedom. However, since the sum of the n_i's is N, there is some dependence among the R_i's. Kruskal (1952) showed that if the ith term in T is multiplied by $(N - n_i)/N$ for $i = 1, 2, \ldots K$, then the result

$$T = \sum_{i=1}^{K} \frac{\{R_i - [n_i(N + 1)/2]\}^2}{n_i(N + 1)N/12}$$

is asymptotically distributed as a chi-square random variable with $K - 1$ degrees of freedom. Since $\sum_{i=1}^{K} R_i = N(N + 1)/2$, T may be written as

$$T = \sum_{i=1}^{K} \frac{\{R_i - [n_i(N + 1)/2]\}^2}{n_i(N + 1)N/12}$$

$$= \frac{12}{N(N + 1)} \sum_{i=1}^{K} \frac{1}{n_i} \left[R_i^2 - R_i n_i(N + 1) + \frac{1}{4} n_i^2(N + 1)^2\right]$$

$$= \frac{12}{N(N+1)} \sum_{i=1}^{K} \frac{R_i^2}{n_i} - \frac{12}{N(N+1)} \left[\frac{N(N+1)}{2} \cdot (N+1) - \frac{N}{4}(N+1)^2\right]$$

$$= \frac{12}{N(N+1)} \sum_{i=1}^{K} \frac{R_i^2}{n_i} - 3(N+1),$$

is an equivalent form for T, and is usually more convenient to use. A modification proposed by Ott uses T' rather than T when there are groups of tied ranks. To do this we form the g groups composed of identical ranks, where the jth group contains t_j ($j = 1, \ldots, g$) ties. The statistic T' is then close to a chi-square random variables with K - 1 degrees of freedom where

$$T' = \frac{T}{1 - [\sum_{j=1}^{g} (t_j^3 - t_j)/N^3 - N)]}.$$

A Generalized Kruskal-Wallis Test for Comparing K Censored Samples

Although the Kruskal-Wallis test assumes only continuous underlying distributions, it does not do very well if there are large numbers of ties. This is especially so for censored data when the ties may lie among the upper values of the ranks.

To handle problems of right censored data, Breslow (1970) generalized the Kruskal-Wallis test. Let X_{ij}° be the true observation for the jth individual obtained from the ith population ($j = 1,\ldots,N_i$; $i = 1,\ldots,K$). Variable Z_{ij} is used to censor X_{ij}°, so sometimes the true observation X_{ij}° may not be observed. The observed data which we can get from a real sample is $X_{ij} = \min(X_{ij}^{\circ}, Z_{ij})$. X_{ij} should indicate with a variable δ_{ij} whether or not X_{ij} is in fact censored: i.e., $\delta_{ij} = 1$ when $X_{ij} = X_{ij}^{\circ} < Z_{ij}$ (uncensored); $\delta_{ij} = 0$ when $X_{ij} = Z_{ij} < X_{ij}$

(censored). $N = N_1 + \ldots + N_k$ is the total sample size and $\lambda_1 = N_i/N$ is the proportion of the ith sample size to the total sample size.

F_i is the ith cumulative distribution function. The null hypothesis to be tested is H_o: $F_1 = \ldots = F_k$, which specified that K populations have equal distribution functions.

Breslow (1970) defined a scoring function χ for comparing two observations X_{ij} and $X_{i'j'}$ by

$$\chi(X_{ij}, \delta_{ij}; X_{i;j;}, \delta_{i'j'}) = \begin{cases} -1 & X_{ij} < X_{i'j'};\ \delta_{ij} = 1,\ \delta_{i'j'} = 1 \\ -1 & X_{ij} < X_{i'j'};\ \delta_{ij} = 1,\ \delta_{i'j'} = 0 \\ +1 & X_{ij} > X_{i'j'};\ \delta_{ij} = 1,\ \delta_{i'j'} = 1 \\ +1 & X_{ij} > X_{i'j'};\ \delta_{ij} = 0,\ \delta_{i'j'} = 1 \\ 0 & \text{otherwise.} \end{cases}$$

The χ function is then used in computing a vector score statistic, $\underline{W}$. The ith component of this vector score statistic is defined to be the total score comparing the ith sample with the remaining K - 1 samples,

$$W_i = \sum_{j=1}^{N_i} \sum_{i'=1}^{K} \sum_{j'=1}^{N_i} \chi(X_{ij}, \delta_{ij}, X_{i'j'}, \delta_{i'j'}).$$

For uncensored data sets, $W_i = 2[R_i - (1/2)N_i(N + 1)]$. Large negative values of W_i mean that observations in the ith sample are smaller than those from other samples and large positive values of W_i would indicate that the ith sample had larger than average values. The total of W_i should be equal to 0.

Breslow (1970) goes on to use this $\underline{W}$ vector to form test statistics for testing the equality of K distribution functions. His first statistic refers to Rao (1965) which shows that the well-known large sample theory for chi-square statistics holds for the statistic

$$S^* = \sum_{i'=1}^{K} \sum_{j'=1}^{N_i} \delta_{i'j'} \{ \sum_{i=1}^{K} \sum_{j=1}^{N_i} e(X_{ij} - X_{i'j'}) \}^2$$

where

$$e(x) = \begin{cases} 1 \text{ if } x > 0 \\ 0 \text{ if } x \leq 0. \end{cases}$$

Under general regularity conditions, S^* can be shown to have an asymptotic chi-square distribution with $K - 1$ degrees of freedom. Breslow recommends evaluation of S^{**} in order to check on computational accuracy of S^*. S^{**} is a lower bound for S^* and is easily computed as:

$$S^{**} = 3N^{-2} \sum_{i=1}^{K} (W_i^2 / N_i).$$

Breslow (1970) goes on to develop a statistic which is calculated as follows. A covariance matrix $\ddagger$ must be computed. Individual terms $\sigma_{ii'}$ can be calculated from

$$N^3 \sigma_{ii'} = - \sum_{i'=1}^{K} \sum_{j'=1}^{N_{i''}} \delta_{i''j''} \sum_{j=1}^{N_{i''}} e(X_{ij} - X_{i''j''}) \sum_{j=1}^{N_{i'}} e(X_{i'j'} - X_{i''j''})$$

$$\sigma_{ii} = - \sum_{i' \neq i} \sigma_{ii'}.$$

where $e(x) = 1$ if $x > 0$, 0 if $x \leq 0$.

The covariance matrix $\ddagger$ is then decomposed into $K - 1$ vectors $\underline{\xi}_i = (\xi_{ii}, \ldots, \xi_{ik})'$ $(i=1, \ldots, K-1)$ such that $\underline{\xi}_i' \ddagger \underline{\xi}_i = 1$ and $\underline{\xi}_i' \ddagger \underline{\xi}_j = 0$ $(i \neq j)$. The vectors $\underline{\xi}_i$ may be easily found by using the Gram-Schmidt orthogonalization process. We can use the $K - 1$ vectors

$\underline{Y}_1 = (1, 0, \ldots, 0)'$, $Y_2 = (1, 1, 0, \ldots, 0)'$, ...,

$$\underline{Y}_{K-1} = (1, \ldots, 1, 0)'$$

as a starting point for the Gram-Schmidt process. We denote the inner product of two vectors by

$$(\underline{X}, \underline{Y}) = \underline{X}' \, \ddagger \, \underline{Y}.$$

The Gram-Schmidt process proceeds as follows:

1. Calculate $\underline{\xi}_1 = \dfrac{\underline{Y}_1}{\|\underline{Y}_1\|}$ where $\|\underline{Y}_1\| = \sqrt{(\underline{Y}_1, \underline{Y}_1)} = \sqrt{\underline{Y}_1' \, \ddagger \, \underline{Y}_1}$

2. Use $\underline{\xi}_1$ to calculate $\underline{Z}_2 = \underline{Y}_2 - (\underline{Y}_2, \underline{\xi}_1)\underline{\xi}_1$ and $\underline{\xi}_2 = \dfrac{\underline{Z}_2}{\|\underline{Z}_2\|}$

   where $\|\underline{Z}_2\| = \sqrt{(\underline{Z}_2, \underline{Z}_2)} = \sqrt{\underline{Z}_2' \, \ddagger \, \underline{Z}_2}$

3. Use above information to calculate $\underline{Z}_3 = \underline{Y}_3 - (\underline{Y}_3, \underline{\xi}_1)\underline{\xi}_1 - (\underline{Y}_3, \underline{\xi}_2)\underline{\xi}_2$

   and $\underline{\xi}_3 = \dfrac{\underline{Z}_3}{\|\underline{Z}_3\|}$ where $\|\underline{Z}_3\| = \sqrt{(\underline{Z}_3, \underline{Z}_3)} = \sqrt{\underline{Z}_3' \, \ddagger \, \underline{Z}_3}$

(K - 1)st. Step. Use above information to calculate

$$\underline{Z}_{K-1} = \underline{Y}_{K-1} - (\underline{Y}_{K-1}, \underline{\xi}_1)\underline{\xi}_1 - (\underline{Y}_{K-1}, \underline{\xi}_2)\underline{\xi}_2 - \cdots$$

$$- (\underline{Y}_{K-1}, \underline{\xi}_{K-2})\underline{\xi}_{K-2} \quad \text{and}$$

$$\underline{\xi}_{K-1} = \frac{\underline{Z}_{K-1}}{\|\underline{Z}_{K-1}\|} \quad \text{where} \quad \|\underline{Z}_{K-1}\| = \sqrt{(\underline{Z}_{K-1}, \underline{Z}_{K-1})} =$$

$$\sqrt{\underline{Z}_{K-1}' \, \ddagger \, \underline{Z}_{K-1}}\,.$$

The statistics

$$S_i = N^{-3/2} \, \underline{\xi}_i' \, \underline{W}$$

are easily found and a combined test statistic is calculated as

$$S = \sum_{i=1}^{K-1} S_i^2$$

which is a chi-square random variable with K - 1 degrees of freedom.

Breslow suggests calculating S** as a lower bound to S*. S and S* are asymptotically equivalent statistics, but S* is computationally far easier to compute. The "easier" S* and S** are needed only if a computer program is not available to calculate S. A computer program is given in the Appendix which translates Breslow's formulae into a FORTRAN IV program for the Burroughs B6800 computer.

CHAPTER III

MULTIPLE COMPARISONS

The procedures described in Chapter II provide an overall test of the equality of $K \geq 2$ distributions. For $K > 2$, if the populations are declared significantly different, then a multiple comparison procedure is needed to isolate the differences.

The Bonferroni inequality provides one method of simultaneously estimating several confidence intervals. Let A_1 denote the first event, say a $1 - \alpha_1$ confidence interval, and let A_2 denote the second event also a $1 - \alpha_2$ confidence interval. We can then use the Bonferroni inequality to get the probability of both events of A_1 and A_2 occurring simultaneously. We already know that

$$P(A_1 \cap A_2) = 1 - P(\overline{A}_1) - P(\overline{A}_2) + P(\overline{A}_1 \cap \overline{A}_2)$$

and since $P(\overline{A}_1 \cap \overline{A}_2) \geq 0$, we obtain the Bonferonni inequality:

$$P(A_1 \cap A_2) \geq 1 - P(\overline{A}_1) - P(\overline{A}_2).$$

For this situation, the joint confidence is

$$P(A_1 \cap A_2) \geq 1 - \alpha_1 - \alpha_2.$$

The Bonferroni inequality can easily be extended to K simultaneous confidence intervals with family confidence coefficient $1 - \alpha$ by requiring $P[\overline{A}_i] = \alpha_i$; and $\Sigma\alpha_i = \alpha$ which then gives

$$P(\bigcap_{i=1}^{K} A_i) \geq 1 - \alpha.$$

For example, let

A_{12} be a 99% confidence interval for $\mu_1 - \mu_2$,

A_{13} be a 99% confidence interval for $\mu_1 - \mu_3$,

and A_{23} be a 99% confidence interval for $\mu_2 - \mu_3$.

The Bonferroni inequality then guarantees us a family or simultaneous confidence interval of at least 97 percent that the three intervals based on the same sample are simultaneously correct, i.e.,

$$P(A_{12} \cap A_{13} \cap A_{23}) \geq .97.$$

If K interval estimates are desired with a family confidence coefficient $1 - \alpha$, constructing each interval estimate with statement confidence coefficient $1 - \alpha/K$ will suffice. The Bonferroni technique is ordinarily most useful when the number of simultaneous estimates is not too large. Note that different statement confidence coefficients also could be calculated, as long as $\sum_{i=1}^{K} P(\bar{A}_i) = \alpha$. For instance, the event A_1 may be a 98 percent confidence interval and the event A_2 could be a 97 percent confidence interval. The family confidence coefficient would then be at least 95 percent.

CHAPTER IV

EXAMPLES

The Methods Used

In this chapter two examples are given to illustrate the analysis of censored data. We will apply the F test for exponential data, the Kruskal-Wallis test for ranked data and Breslow's method for censored data. The Bonferroni method will then be used to test pairwise comparisons.

Example 1

In the first example, three groups of data were generated from known exponential distributions. The procedure to get the three data sets uses an integral transform, i.e., if $F(X)$ is the distribution function for a random variable X, and if $X_1, \ldots, X_n$ is a random sample from $F(\cdot)$ then $U_i = F(X_i)$ for $i = 1, \ldots, n$ will be a random sample of uniform random variables over the interval (0, 1). It follows then that if $U_1, \ldots, U_n$ is a random sample from a uniform distribution, then $X_i = F^{-1}(U_i)$ for $i = 1, \ldots, n$ will be a random sample from $F(\cdot)$.

It is easy to use a computer to generate uniform random numbers, and then we may use the integral transformation technique for finding random numbers from a given distribution. For this example we got three groups of uniform random numbers from 0 to 1 using MINITAB. If $F(\cdot)$ is a negative exponential distribution, $F_X(X) = 1 - e^{-X/\lambda} = \mu$, then $X = F_X^{-1}(\mu) = -\lambda \ln (1 - \mu)$ has a negative exponential distribution with parameter λ. i.e., the density function of X is $f_X(X) = (1/\lambda)e^{-X/\lambda}$, which is a negative exponential distribution. Table 1 presents such samples from three negative exponential distributions. Each sample has been sorted for ease in censoring at an arbitrary value of 20.

If we ignore the fact that the data is censored, we can get $\hat{\lambda}_i$ ($\hat{\lambda}_i = X_{i.}/\delta_{i.}$) for each group and then use $R = \hat{\lambda}_i/\hat{\lambda}_j$ ($i \neq j$) to do an F test with $2\delta_{i.}$ and $2\delta_{j.}$ degree of freedom. All possible F tests are listed in Table 2.

If we ignore the censoring in Table 1, Table 3 then gives the R_i's, the sum of the ranks for each group, and the Kruskal-Wallis tests are listed in Table 4.

Since Breslow's method includes many complicated formulae, the computer program listed in the Appendix was used to get the statistics S**, S* and S. We used S to do the chi-square test and calculated S* and S** for illustrative purpose only. Using an experimentwise error rate of .05, the three pairwise comparisons are $G_1 = G_2$, $G_1 = G_3$ and $G_2 = G_3$. The Bonferroni procedure then uses 1 - .05/3 = .9833 as confidence coefficient for the individual intervals. We list the results of the F test, the Kruskal-Wallis test, Breslow's method and the Bonferroni method in Table 5.

Table 1. Data and λ_i for sorted Monte Carlo data.

Treatment	Group 1	Group 2	Group 3
True λ_i	12	10	5
n_i	15	12	18
Sorted Uncensored Data	1.6162	0.3113	1.0977
	3.3108	0.3193	1.5408
	6.4986	0.5793	1.7251
	8.7001	0.9393	2.1613
	9.5813	1.4585	2.8869
	10.3311	2.6174	3.0430
	11.3729	2.8732	3.1428
	11.6698	6.5377	3.5553
	14.7173	12.9082	3.9252
	20.1485	20.1756	4.2688
	28.3848	22.2597	4.3705
	32.3451	38.9166	6.9837
	34.7911		7.7971
	51.9654		7.9731
	66.7175		8.2069
			9.2274
			11.2849
			20.2483
$\overline{X}$ (Uncensored)	20.81	9.158	5.7466
$\delta_{i.} = \sum_{j=1}^{n_i} \delta_{ij}$	9	9	17
$X_{i.} = \sum_{j=1}^{n_i} X_{ij}$ for data censored at 20	197.798	88.544	103.191
$\hat{\lambda}_i = X_{i.}/\delta_{i.}$	21.978	9.838	6.07

Table 2. The results of example 1 using F test for censored data in Table 1.

Hypothesis	$G_1 = G_2$	$G_1 = G_3$	$G_2 = G_3$
$R = \hat{\lambda}_i/\hat{\lambda}_j$	2.234[a]	3.621[a]	1.621
Degrees of Freedom	18, 18	18, 34	18, 34
P Value[b]	0.04845	0.00060	0.11002

[a]Significant at $\alpha = 0.05$.

[b]Run STATPAC/DIST to get the probability of an F value larger than observed when the degrees of freedom are $2\delta_{i.}$ and $2\delta_{j.}$ respectively.

Table 3. The ranks for data censored at 20 from Table 1.

Hypothesis	$G_1 = G_2$		$G_1 = G_3$		$G_2 = G_3$		$G_1 = G_2 = G_3$		
Groups	G_1	G_2	G_1	G_3	G_2	G_3	G_1	G_2	G_3
$R(X_{ij})$	6	1	3	1	1	5	8	1	5
	9	2	9	2	2	7	16	2	7
	10	3	14	4	3	8	21	3	9
	12	4	19	5	4	9	27	4	10
	13	5	21	6	6	12	29	6	13
	14	7	22	7	10	13	30	11	14
	15	8	24	8	11	14	32	12	15
	16	11	25	10	19	15	33	22	17
	18	17	26	11	26	6	35	34	18
	23	23	30	12	28.5	17	40.5	40.5	19
	23	23	30	13	28.5	18	40.5	40.5	20
	23	23	30	15	28.5	20	40.5	40.5	23
	23		30	16		21	40.5		24
	23		30	17		22	40.5		25
	23		30	18		23	40.5		26
				20		24			28
				23		25			31
				30		28.5			40.5
$R_i = \sum_{j=1}^{n_i} R(X_{ij})$	251	127	343	218	167.5	297.5	474	216.5	344.5

Table 4. The results of example 1 using Kruskal-Wallis test and the ranks given in Table 3.

Hypothesis	$G_1 = G_2$	$G_1 = G_3$	$G_2 = G_3$	$G_1 = G_2 = G_3$
T	4.0024	10.1229	0.6134	9.70
T'	4.1546[a]	10.2185[a]	0.6148	9.8066[a]
P Value[b]	0.04152	0.00139	0.43299	0.00742

[a]Significant at $\alpha = .05$.

[b]Probability of a X^2 value larger than observed.

Table 5. Statistics S**, S* and S, and the results of example 1 for F test, Kruskal-Wallis test, Breslow's method and Bonferroni method.

Method	Hypothesis	$G_1 = G_2$	$G_1 = G_3$	$G_2 = G_3$	$G_1 = G_2 = G_3$
Breslow and Bonferroni	S**	4.1506	10.4296	0.6338	9.9127
	S*	4.5410	11.0086	0.6679	10.3524
	S	5.0971[a]	11.7467[a,b]	0.7309	9.7876[a]
Kruskal-Wallis and Bonferroni	T'	4.1546[a]	10.2185[a,b]	0.6148	9.8066[a]
	χ^2_{K-1}, .95	3.84	3.84	3.84	5.991
F Test	R	2.234[c]	3.621[c]	1.621	
P Values	S	0.02397	0.00061	0.39259	0.00749
	T'	0.04152	0.00139	0.43299	0.00742
	R	0.04845	0.00060	0.11002	———

[a]Significant for chi-square value with $\alpha = .05$.

[b]Significant for chi-square value with $\alpha = .05/3 = .01667$.

[c]Significant for F value with $\alpha = .05$.

From Table 5, it is easy to see that the statistic S** is a lower bound to S* and it also is a lower bound to S except for the case $G_1 = G_2 = G_3$. T', the tie-corrected Kruskal-Wallis test statistic, is very close to S** in this example. To allow easy comparison of the S, T' and R test results, p-values were obtained by running STATPAC/DIST to get the probability of a large chi-square or F statistic. For this data set the statistic R always got the smallest probability except for the test of $G_1 = G_2$. This means that when we test the null hypothesis of equality of two groups, the R value is possibly too liberal, i.e., too easy to reject. The probability of S and T' listed in Table 5 show that the p-value for S is always smaller than the p-value of T' except the case $G_1 = G_2 = G_3$. In this example, the three statistics S, T' and R yielded the same conclusions, i.e., the "significant" difference between group 1 and groups 2 and 3. These results are somewhat surprising since there is a relatively small difference between the λ_i's for groups 1 and 2. Since this was Monte Carlo data, these differences may be ascribed to chance. Further Monte Carlo work would undoubtedly tend to "smooth" these unexpected differences.

Example 2

The second example is a nutrition experiment conducted by Susan Collinge who was a graduate student in Nutrition Food Science Department, USU, in 1981. Susan looked at how many samples of meat products were bad each day when they were put in 27° C (80.6° F) temperature room. Each of the nine treatments contained twenty-five sealed bags of meat with different chemical additives. During each day a count of the number of swollen bags was made. The swelling indicated spoilage of the contents. After 100 days the experiment was terminated, resulting in some treatments having censored data. The results of the nine treatments are compared below to see what kind of chemical combination added to the meat will keep the meat from spoilage for the longest period of time. The nine treatments were:

Treatment 1. Control - no chemicals

Treatment 2. Nitrite only

Treatment 3. Nitrite + 20 ppm $FeCl_3$

Treatment 4. Nitrite + Myoglobin

Treatment 5. Nitrite + 200 ppm EDTA + Myoglobin

Treatment 6. Nitrite + Nytrosylmyoglobin

Treatment 7. Nitrite + 200 ppm EDTA

Treatment 8. Nitrite + 200 ppm EDTA + 20 ppm $FeCl_3$

Treatment 9. Nitrite + 200 ppm EDTA + 40 ppm $FeCl_3$

We are interested in the following specific comparisons:

1. Treatment 1 vs. treatment 2 through 9.
2. Treatment 2 vs. treatment 3 through 9.
3. Treatment 3 vs. treatment 4.
4. Treatment 3 vs. treatment 8.
5. Treatment 4 vs. treatment 6.
6. Treatment 5 vs. treatment 7.
7. Treatment 7 vs. treatment 8.
8. Treatment 8 vs. treatment 9.

Table 6 gives the values n_i, $\delta_{i\cdot}$, $X_{i\cdot}$ and $\hat{\lambda}_i$, and the results of the F test are listed in Table 7. The R_i's and results of the Kruskal-Wallis test are shown in Table 8. Table 9 shows all the results of the

F test, the Kruskal-Wallis test, Breslow's method and the Bonferroni method. From Table 9 we find the different methods yield the same results, i.e., only treatment 3 and treatment 4 are homogeneous.

Table 6. The values of n_i, $\delta_{i.}$, $X_{i..}$ and $\hat{\lambda}_i$ of each treatment.

Treatment	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	$T_2 - T_9$	$T_3 - T_9$
n_i	25	25	25	25	25	25	25	25	25	200	175
$\delta_{i.}$	25	25	23	25	14	13	1	11	1	113	88
$X_{i.}$	100	559	587	473	1457	1893	2424	1806	2421	11620	11061
$\hat{\lambda}_i = X_{i.}/\delta_{i.}$	4	22.36	25.52	18.92	104.071	145.615	2424	164.182	2421	102.83	125.69

Table 7. The results for example 2 using F test.

Hypothesis	T_1 vs $T_2 - T_9$	T_2 vs $T_3 - T_9$	T_3 vs T_4	T_3 vs T_8	T_4 vs T_6	T_5 vs T_7	T_7 vs T_8	T_8 vs T_9
$R = \hat{\lambda}_i/\hat{\lambda}_j$	25.7075[a]	5.621[a]	1.3488	6.433[a]	7.696[a]	23.292[a]	14.764[a]	14.746[a]
Degrees of Freedom	226,50	175,50	46,50	22,46	26,50	2,28	2,22	2,22
P Values[b]	< .00001	< .00001	0.15046	0.00001	< .00001	0.00001	0.00009	0.00009

[a] Significant at $\alpha = .05$.

[b] RUN STATPAC/DIST to get the p values.

Table 8. The results of example 2 using Kruskal-Wallis test.

Hypothesis	T_1 vs $T_2 - T_9$	T_2 vs $T_3 - T_9$	T_3 vs T_4	T_3 vs T_8	T_4 vs T_6	T_5 vs T_7	T_7 vs T_8	T_8 vs T_9
$N = \Sigma n_i$	225	200	50	50	50	50	50	50
n_i	25 200	25 175	25 25	25 25	25 25	25 25	25 25	25 25
R_i	325 25100	1143.5 18956.5	584.5 690.5	399.5 875.5	364.5 910.5	476.5 798.5	761 514	514 761
T	66.3717	25.5749	1.0575	21.3248	28.0580	9.7585	5.7420	5.7420
T'	70.5512[a]	27.8846[a]	1.0617	22.1017[a]	28.4833[a]	14.8667[a]	10.2335[a]	10.2335[a]
P Values for T'	< .00001	0.00001	0.30283	0.00001	0.00001	0.00012	0.00138	0.00138

[a]Significant at $\alpha = 0.05$.

Table 9. The results of example 2 using F test, Kruskal-Wallis test, Breslow's method and Bonferroni method.

Method	Hypothesis	T_1 vs $T_2 - T_5$	T_2 vs $T_3 - T_9$	T_3 vs T_4	T_3 vs T_8	T_4 vs T_6	T_5 vs T_7	T_7 vs T_8	T_8 vs T_9
Breslow and Bonferroni	S**	66.6667	25.7028	1.0787	21.7513	28.6191	9.9537	5.8569	5.8569
	S*	74.8929	29.0374	1.2002	24.4947[a]	30.8067[a]	16.0582[a]	10.7245[a]	10.7245[a]
	S	2659.5745[a]	49.5521[a]	1.2317	28.6045[b]	38.3626[b]	16.6990[b]	10.9237[b]	10.9237[b]
Kruskal-Wallis and Bonferroni	T'	70.5512[a]	27.8846[a]	1.0617	22.1017[a,b]	28.4833[a,b]	14.8667[a,b]	10.2335[a,b]	10.2335[a,b]
F Test	R	25.7075[c]	5.621[c]	1.3488	6.433[c]	7.696[c]	23.292[c]	14.764[c]	14.746[c]
P Values	S	< .00001	< .00001	0.26708	0.00001	< .00001	0.00005	0.00095	0.00095
	T'	< .00001	0.00001	0.30283	0.00001	0.00001	0.00012	0.00138	0.00138
	R	< .00001	< .00001	0.15046	0.00001	< .00001	0.00001	0.00009	0.00009

[a]Means significant for chi-square value with $\alpha = .05$.

[b]Means significant for chi-square value with $\alpha = .05/6 = .00833$.

[c]Means significant for F value with $\alpha = .05$.

From Table 9 we see that for this example S** was a lower bound to S* and S. The statistic T' was between S** and S*, sometimes it was close to S** and sometimes close to S*. In this example the p value for R was the smallest value for all of the cases, whereas the T' value of Kruskal-Wallis test had the biggest probability. This suggests that the R test may be too liberal (too easy to reject) while the Kruskal-Wallis test (corrected for ties) may be too conservative. Since the true population values are unknown for this case, it is impossible to say for certain.

CHAPTER V

CONCLUSIONS

A fundamental problem in many biological and medical investigations is a comparison of the survival distributions from two or more samples of censored data. The hypothesis of interest is the equality of survival time distribution functions across samples. In this paper we discussed this topic and analyzed censored data by using four different methods, namely: the F test, the Kruskal-Wallis test, Breslow's generalization of the Kruskal-Wallis test and a Bonferroni multiple comparison method. The F test is restricted to two exponential distributions, so it cannot be used widely. The Kruskal-Wallis test is suitable for two or more populations, but for censored data there will usually be a lot of ties. This violates the assumptions made in developing the Kruskal-Wallis test. For the Bonferroni method we use a given value α to do the K multiple comparisons, then for each single case will only use $1 - \alpha/K$ to test the hypothesis. In this situation the given confidence interval is so large that it is hard to reject the null hypothesis. If a computer is available, we can translate the formulae of Breslow's method to a computer program as given in the Appendix. It will then be easy to analyze censored data. For all the reasons stated above, we prefer to use Breslow's method if a computer is available. If not, the Kruskal-Wallis procedure corrected for ties seemed to give almost the same results. There are only two examples in this paper; if we want to get more information to tell the exact differences between the Kruskal-Wallis test and Breslow's method, more examples and a computer program for the Kruskal-Wallis test should be developed.

A Monte Carlo study could give enough different situations involving different distributions and different values of the parameters to help decide on the best overall procedure. A more exact multiple comparison procedure could also be developed using the asymptotical distribution of Breslow's vector of W_i's rather than using the Bonferroni inequality. Since the Bonferroni method seems to work fairly well in these examples, further work might not be terribly worthwhile. Further Monte Carlo research could help in the decision on whether to puruse this matter in more detail.

REFERENCES

Breslow, N. 1970. A generalized Kruskal-Wallis test for comparing K samples subject to unequal patterns of censorship. Biometrika 57:579-94.

Conover, W.J. 1971. Practical Nonparametric Statistics. New York: John Wiley and Sons, Inc.

Lagakos, S.W. 1979. General right censoring and its impact on the analysis of survival data. Biometrics 35:139-156.

Kruskal, W.H. 1952. A nonparametric test for the several sample problem. The Annals of Mathematical Statistics 23:525-540.

Kruskal, W.H. and Wallis, W.A. 1952. Use of ranks in one-criterion variance analysis. Journal of the American Statistical Association 47:583-621.

Mood, Alexander M., Graybill, Franklin A., and Boes, Duane C. 1974. Introduction to the Theory of Statistics. New York: McGraw Hill, Inc.

Neter, John and Wasserman, William. 1974. Applied Linear Statistical Models. Homewood, Illinois: Richard D. Irwin, Inc.

Ott, Lyman. 1977. An Introduction to Statistical Method and Data Analysis. Belmont, CA: Wadsworth Publishing Company, Inc.

Wallis, W.A. 1939. The correlation ratio for ranked data. Journal of the American Statistical Association 34:533-538.

APPENDIXES

```
$SET  LINEINFO AUTOBIND
$BIND=FROM IMSL/=
C* THE IMSL(INTERANTIONAL MATHEMATICAL AND STATISTICAL
C* LIBRARIES) LIBRARY CONSISTS  OF A SUBSTANTIAL COLLECTION
C* OF SUBROUTINES AND FUNCTIONS SUBPROGRAMS IN THE AREAS OF
C* MATHEMATICS AND STATISTICS.
FILE  5(KIND=DISK,FILETYPE=7)
C*
C*      ***************************************************
C*      *                                                 *
C*      * BIOMETRIKA (1970),57,3,P,579                    *
C*      * A GENERALIZED KRUSKAL-WALLIS TEST FOR COMPARING *
C*      * K SAMPLES SUBJECT TO UNEQUAL PATTERNS OF        *
C*      * CENSORSHIP.     BY NORMAN BRESLOW               *
C*      * FORTRAN PROGRAM WRITTEN FOR THE BURROUGHS B6800 *
C*      * COMPUTER AT USU BY HWEI-WENG SHIH IN PARTIAL    *
C*      * FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE  *
C*      * OF MASTER OF SCIENCE IN APPLIED STATISTICS AT   *
C*      * UTAH STATE UNIVERSITY,1981.                     *
C*      *                                                 *
C*      ***************************************************
C*
      DIMENSION N(10),W(10),X(10,200),D(10,200),SIGE(10,10),
     *          SIG(10,10),W1(10),DL(10),YY(10,10),XX(10,10),
     *          Y1(10),Y2(10),X1(10),ANS(10),V(10),S(10),
     *          DX(10,200),DD(10,200),ND(10),INDX(2)
C* THE DATA ARE ENTERED TO THE DATA FILE BY FOLLOWING STEPS:
C*    1. ENTER K, K IS THE NUMBER OF SAMPLES
C*    2. ENTER N(1), N(1) IS THE NUMBER OF OBSERVATIONS
C*       OF THE FIRST SAMPLE
C*    3. ENTER PAIRS DATA X(1,1),D(1,1);X(1,2),D(1,2);
C*       ...;X(1,N(1)),D(1,N(1))
C*    4. ENTER N(2), N(2) IS THE NUMBER OF OBSERVATIONS
C*       OF THE SECOND SAMPLE
C*    5. ENTER PAIRS DATA X(2,1),D(2,1);X(2,2),D(2,2);
C*       ...;X(2,N(2)),D(2,N(2))
C*    .
C*    .
C*    .
C*    2K. ENTER N(K), N(K) IS THE NUMBER OF OBSERVATIONS
C*        OF THE LAST SAMPLE
C*    2K+1. ENTER PAIRS DATA X(K,1),D(K,1);X(K,2),D(K,2);
C*          ...;X(K,N(K)),D(K,N(K))
      READ(5,/) K
      NT=0
      DO 50 I=1,K
      READ(5,/) N(I)
      NT=NT+N(I)
      READ(5,/) (X(I,J),D(I,J),J=1,N(I))
   50 CONTINUE
C***
C**   DENOTE BY XO(I,J) THE TRUE OBSERVATION FOR THE (J)TH
C**   INDIVIDUAL IN THE (I)TH SAMPLE (J=1,...,N(I),I=1,...,K),
C**   SINCE THIS OBSERVATION MAY BE CENSORED BY A VARIABLE
C**   Z(I,J), IT CANNOT ALWAYS BE OBSERVED, RATHER ONE
C**   OBSERVES
C**        X(I,J)=MIN(XO(I,J),Z(I,J))
C**   ALONG WITH THE INDICATOR VARIABLE
```

```
C**        D(I,J)=1  IF X(I,J)=XO(I,J)
C**        D(I,J)=0  IF X(I,J)=Z(I,J)<XO(I,J)
C**    N(I)=# OBSERVATION OF EACH SAMPLE
C***
       NT1=NT**3
       WRITE(6,101)
       WRITE(6,100) NT,NT1
  100  FORMAT(1X,' NT=THE TOTAL SAMPLE SIZE=',I4,/,
     *        ' NT1=THE CUBIC OF TOTAL SAMPLE SIZE=',I10)
  101  FORMAT(1X,///,50(1H*),///)
C***
C**    P1 IS THE FIRST SUBROUTINE AND IT DEFINE  A SCORING
C**    FUNCTION PSI, THEN COMPUTE THE VECTOR W(I).
C***
       CALL P1(K,N,W,X,D)
C***
C**    P2 IS THE SECOND SUBROUTINE AND IT COMPUTE  THE
C**    INDIVIDUAL TERMS OF A COVARIANCE MATRIX SIGMA.
C***
       CALL P2(K,N,NT1,W,X,D,SIGE,SIG)
C***
C**    P3 IS THE THIRD SUBROUTINE AND IT COMPUTE THE
C**    STATISTICS S* AND S**.
C***
       CALL P3(K,N,NT,NT1,W,W1,DL,X,D,S3,S2)
C***
C**    P4 IS THE FOURTH SUBROUTINE AND IT COMPUTE THE
C**    STATISTIC S.
C***
       CALL P4(K,NT1,YY,XX,SIGE,X1,Y1,Y2,W,ANS,V,S,TOT)
       WRITE(6,111)
  111  FORMAT(1X,///,' ----- TESTING HYPOTHESIS -----')
C* MDCH IS AN IMSL SUBROUTINE WHICH IS USED TO GET
C* THE PROBABILITY OF A CHI-SQUARE DISTRIBUTION.
C* THE FORM IS: CALL KMDCH(CS,DF,P,IER)
C* CS - INPUTE VALUE FOR WHICH THE PROBABILITY IS
C*       COMPUTED.
C* DF - INPUTE NUBER OF DEGREES OF FREEDOM OF THE
C*       CHI-SQUARE DISTRIBUTION.
C* P - OUTPUT PROBABILITY THAT A RANDOM VARIABLE
C*      WHICH FOLLOWS THE CHI-SQUARE DISTRIBUTION
C*      WITH DF DEGREES OF FREEDOM IS LESS THAN OR
C*      EQUAL TO CS.
C* IER - ERROR  PARAMETER.
       CALL MDCH(S3,K-1,PS3,IER)
       PS3=1-PS3
       WRITE(6,102) S3
  102  FORMAT(1X,/,' S**=',F11.6)
       WRITE(6,103) K-1,S3,PS3
  103  FORMAT(1X,/,2X,' P(CHI-SQUARE(',I2,') >=',F11.6,' ) =',F8,6)
       CALL MDCH(S2,K-1,PS2,IER)
       PS2=1-PS2
       WRITE(6,104)S2
  104  FORMAT(1X,//,'  S*=',F11.6)
       WRITE(6,103) K-1,S2,PS2
       CALL MDCH(TOT,K-1,PTOT,IER)
       PTOT=1-PTOT
       WRITE(6,105) TOT
  105  FORMAT(1X,//,'  S=',F11.6)
       WRITE(6,103) K-1,TOT,PTOT
       WRITE(6,106)
```

```
  106 FORMAT(1X,//,' S** AS A LOWER BOUND TO S*, S AND S* WILL BE
     *          ',/,' ASYMPTOTICALLY EQUIVALENT STATISTICS,'
     *          ,/,' S* IS COMPUTATIONALLY SIMPLER THAN S,'
     *          ,/,' HOWEVER, ONLY S WILL BE AN ASYMPTOTICALLY'
     *          ,/,' VALID STATISTIC UNDER HYPOTHESIS.')
      WRITE(6,101)
      IF(K.LT.3) GO TO 500
C***
C**   THIS PART USE BONFERRONI MULTIPLE COMPARISON METHOD
C**   NPWC=NUMBER OF PAIRWISE COMPARISON
C**   1-ALPHA=CONFIDENT COEFFICIENT
C**   INDX(1) AND INDX(2) ARE THE TWO GROUPS WHICH
C**   WANT TO COMPAIR
C***
      READ(5,/) NPWC,ALPHA
      DO 300 KK=1,NPWC
      READ(5,/) INDX(1),INDX(2)
      WRITE(6,114)
  114 FORMAT(1X,50(1H*))
      WRITE(6,115) INDX(1),INDX(2)
  115 FORMAT(1X,//,'   ----- ',I1,' VS ',I1,' -----',///)
         DO 200 I=1,2
         ND(I)=N(INDX(I))
         DO 200 J=1,N(INDX(I))
         DD(I,J)=D(INDX(I),J)
  200    DX(I,J)=X(INDX(I),J)
      NT=0
         DO 250 I=1,2
  250    NT=NT+N(INDX(I))
      NT1=NT**3
      WRITE(6,100) NT,NT1
      CALL P1(2,ND,W,DX,DD)
      CALL P2(2,ND,NT1,W,DX,DD,SIGE,SIG)
      CALL P3(2,ND,NT,NT1,W,W1,DL,DX,DD,S3,S2)
      CALL P4(2,NT1,YY,XX,SIGE,X1,Y1,Y2,W,ANS,V,S,TOT)
      WRITE(6,111)
      CALL MDCH(S3,1,PS3,IER)
      PS3=1-PS3
      WRITE(6,102) S3
      WRITE(6,107) S3,PS3
  107 FORMAT(1X,/,2X,' P[CHI-SQUARE(1) >=',F9.6,' ] =',F8.6)
      CALL MDCH(S2,1,PS2,IER)
      PS2=1-PS2
      WRITE(6,104) S2
      WRITE(6,107) S2,PS2
      CALL MDCH(TOT,1,PTOT,IER)
      PTOT=1-PTOT
      TEST=ALPHA/NPWC
      WRITE(6,105) TOT
      WRITE(6,107) TOT,PTOT
      WRITE(6,125) PTOT,PTOT,ALPHA,NPWC,TEST
  125 FORMAT(1X,' USING BONFERRONI INEQUALITY',
     *        /,' REJECT ',F8.6,' IF ',F8.6,' <',F5.2,'/',I1,'=',F8.6)
      WRITE(6,106)
      WRITE(6,108) ALPHA,NPWC,ALPHA,NPWC
  108 FORMAT(1X,///,' BONFERRONI CRITICAL VALUE',
     *        /,' (ASSUMES ALPHA=',F5.2,',',I2,' PAIRWISE COMPARISONS)',
     *        /,' =CHI-SQUARE FOR 1DF, (1 -',F5.2,'/',I1,')100(TH)%')
      P=1-ALPHA/NPWC
C* MDCHI IS AN IMSL SUBROUTINE WHICH IS USED TO GET
C* THE INVERSE VALUE OF A CHI-SQUARE DISTRIBUTION.
```

```
C* THE FORM IS: CALL MDCHI(P,DF,X,IER)
C* P - INPUT PROBABILITY.
C* DF - INPUT NUMBER OF DEGREES OF FREEDOM.
C* X - OUTPUT CHI-SQUARE VALUE, SUCH THAT A RANDOM
C*      VARIABLE, DISTRIBUTED AS CHI-SQUARE WITH DF
C*      DEGREES OF FREEDOM, WILL BE LESS THAN OR
C*      EQUAL TO X WITH PROBABILITY P.
C* IER -   ERROR PARAMETER.
      CALL MDCHI(P,1,CHI,IER)
      WRITE(6,109) CHI,P
  109 FORMAT(1X,/,2X,' P(CHI-SQUARE(1) <=',F9.6' ) =',F8.6)
      WRITE(6,101)
  300 CONTINUE
  500 STOP
      END
C* K - THE NUMBER OF SAMPLES
C* N(I) - THE NUMBER OF OBSERVATIONS OF THE (I)TH SAMPLE?
C* W(I) - THE VECTOR SCORE STATISTIC
C* X(I,J) AND D(I,J) -
C* IF X(I,J) IS AN UNCENSORED DATA THEN D(I,J)=1;
C* IF X(I,J) IS A CENSORED DATA THEN D(I,J)=0.
      SUBROUTINE P1(K,N,W,X,D)
C***
C**   THIS SUBROUTINE DEFINE A SCORING FUNCTION PSI, THEN
C**   COMPUTE THE VECTOR W(I).
C**   WE DEFINE THE SCORING FUNCTION PSI FROM EQUATION (3)
C**   OF BRESLOW FOR COMPARING TWO OBSERVATION X(I,J)
C**   AND X(I',J') BY
C**
C**       PSI=-1  IF X(I,J) < X(I',J'),
C**                  DELTA(I,J)=1, DELTA(I',J')=1
C**       PSI=-1  IF X(I,J) < X(I',J'),
C**                  DELTA(I,J)=1, DELTA(I',J')=0
C**       PSI=+1  IF X(I,J) > X(I',J'),
C**                  DELTA(I,J)=1, DELTA(I',J')=1
C**       PSI=+1  IF X(I,J) > X(I',J'),
C**                  DELTA(I,J)=0, DELTA(I',J')=1
C**       PSI=0   OTHERWISE
C**
C**   THE (I)TH COMPONENT, W(I), OF THE VECTOR SCORE
C**   STATISTIC IS DEFINED TO BE THE TOTAL SCORE
C**   COPARING THE (I)TH SAMPLE WITH THE REMAINING
C**   K-1 SAMPLES.
C**
C**       W(I)=SUM J=1, N(I);
C**                I'=1, K;
C**                J'=1, N(I')
C**             OF PSI(X(I,J), DELTA(I,J);
C**                   X(I',J'),DELTA(I',J'))
C***
      DIMENSION N(10),W(10),X(10,200),D(10,200)
      WRITE(6,140)
  140 FORMAT(1X,///,' ===== VECTOR SCORE STATISTIC =====')
      DO 110 I=1,K
      W(I)=0
         DO 120 J=1,N(I)
         DO 120 IP=1,K
         IF(I.EQ.IP) GO TO 120
            DO 130 JP=1,N(IP)
             T1=D(I,J)+D(IP,JP)
             T2=D(I,J)*D(IP,JP)
```

```
            IF(T1.EQ.0) GO TO 130
            IF(T2.EQ.0) GO TO 10
            IF(X(I,J).LT.X(IP,JP)) GO TO 20
            IF(X(I,J).GT.X(IP,JP)) GO TO 30
            GO TO 130
   10       IF(D(I,J).GT.D(IP,JP)) GO TO 40
            IF(X(I,J).GT.X(IP,JP)) GO TO 30
            GO TO 130
   40       IF(X(I,J).LT.X(IP,JP)) GO TO 20
            GO TO 130
   20       W(I)=W(I)-1
            GO TO 130
   30       W(I)=W(I)+1
  130       CONTINUE
  120     CONTINUE
        WRITE(6,102) I,W(I)
  102 FORMAT(1X,/,' W(',I2,')=',F7.0)
        SUMW=SUMW+W(I)
  110 CONTINUE
        WRITE(6,135) SUMW
  135 FORMAT(1X,/,' THE SUM OF W(I) EQUAL ',F2.0)
        RETURN
        END
C* K - THE NUMBER OF SAMPLES
C* N - THE NUMBER OF OBSERVATIONS OF THE (I)TH SAMPLE
C* NT1 - THE CUBIC OF TOTAL SAMPLE SIZE
C* W(I) - THE VECTOR SCORE STATISTIC
C* X(I,J) AND D(I,J) - SOURCE DATA
        SUBROUTINE P2(K,N,NT1,W,X,D,SIGE,SIG)
C***
C**    THIS SUBROUTINE COMPUTE THE INDIVIDUAL TERMS
C**    OF A COVARIANCE MATRIX SIGMA?
C**    TERMS SIG(I,I'), IN COVARIANCE MATRIX SIGMA
C**    MAY BE FOUND FROM THE FORMULAE OF EQUATION (8)
C**    OF BRESLOW
C**
C**        I UNEQUAL I'
C**        (N**3)*(SIG(I,I'))
C**        =-SUM [I''=1,K; J''=1,N(I'') OF DELTA(I'',J''))
C**              (SUM J=1,N(I) OF E(X(I,J)-X(I'',J'')))
C**              (SUM J'=1,N(I') OF E(X(I',J')-X(I'',J''))))]
C**        E(X)=1  IF X > 0
C**        E(X)=0  IF X < 0, OR X = 0
C**
C**        I EQUAL I'
C**        SIG(I,I)=-SUM I' UNEQUAL I OF SIG(I,I')
C***
        DIMENSION N(10),W(10),X(10,200),D(10,200),SIGE(10,10),
     *            SIG(10,10)
        DO 210 I=1,K
        DO 210 IP=1,K
        IF(I.EQ.IP) GO TO 210
        TOT=0
        SIGE(I,IP)=0
          DO 220 IPP=1,K
          DO 220 JPP=1,N(IPP)
          IF(D(IPP,JPP).EQ.0) GO TO 220
          EPS1=0
            DO 230 J=1,N(I)
            IF(X(I,J).LE.X(IPP,JPP)) GO TO 230
            EPS1=EPS1+1
```

```
 230         CONTINUE
          EPS2=0
             DO 240 JP=1,N(IP)
             IF(X(IP,JP).LE.X(IPP,JPP)) GO TO 240
             EPS2=EPS2+1
 240         CONTINUE
          EPS=EPS1*EPS2
          TOT=TOT+EPS
 220      CONTINUE
       SIGE(I,IP)=-TOT/NT1
 210   CONTINUE
       DO 250 I=1,K
       SIG(I,I)=0
          DO 260 IP=1,K
          IF(IP.EQ.I) GO TO 260
          SIG(I,I)=SIG(I,I)-SIGE(I,IP)
 260      CONTINUE
       SIGE(I,I)=SIG(I,I)
 250   CONTINUE
       WRITE(6,204)
 204   FORMAT(1X,///,' ----- SIGMA MATRIX -----')
       DO 270 I=1,K
 270   WRITE(6,205) (SIGE(I,J),J=1,K)
 205   FORMAT(1X,/,9(1X,F7.4))
 280   RETURN
       END
C* K - THE NUMBER OF SAMPLES
C* N(I) - THE NUMBER OF OBSERVATIONS OF THE (I)TH SAMPLE
C* NT - THE TOTAL OBSERVATIONS
C* NT1 - THE CUBIC OF TOTAL OBSERVATIONS
C* W(I) - THE VECTOR SCORE STATISTIC
C* X(I,J) AND D(I,J) - SOURCE DATA
       SUBROUTINE P3(K,N,NT,NT1,W,W1,DL,X,D,S3,S2)
C***
C**    THIS SUBROUTINE COMPUTE THE STATISTICS S* AND S**,
C**    THE WELL-KNOWN LARGE SAMPLE THEORY FOR CHI-SQUARED
C**    STATISTICS IS FROM EQUATION (12) AND TOP OF P.584
C**    OF BRESLOW
C**
C**        S*=1/[SUM I'=1,K; J'=1,N(I') OF DELTA(I',J')
C**              (SUM I=1,K; J=1,N(I) OF E(X(I,J)-X(I',J')))**2]
C**            *(SUM I=1,K OF W(I)**2/LAMBDA(I))
C**        LAMBDA(I)=N(I)/N
C**
C**    THIS SUGGESTS EVALUATING S** AS A LOWER BOUND TO
C**    S* IN ORDER TO CHECK ON COMPUTATIONS.
C**
C**        S**=(3/N**3)*(SUM I=1,K OF W(I)**2/LAMBDA(I))
C***
       DIMENSION W(10),W1(10),DL(10),X(10,200),D(10,200)
       REAL N(10),NT
       SB=0
       DO 310 IP=1,K
       DO 310 JP=1,N(IP)
       IF(D(IP,JP).EQ.0) GO TO 310
       EPS=0
          DO 320 I=1,K
          DO 320 J=1,N(I)
          IF(X(I,J).LE.X(IP,JP)) GO TO 320
          EPS=EPS+1
 320      CONTINUE
```

```
      SA=D(IP,JP)*EPS**2
      SB=SB+SA
  310 CONTINUE
      SC=0
      DO 330 I=1,K
      DL(I)=N(I)/NT
      W1(I)=W(I)*W(I)
      SC=SC+W1(I)/DL(I)
  330 CONTINUE
      S3=3*SC/NT1
      S2=SC/SB
      RETURN
      END
C* K - THE NUMBER OF SAMPLES
C*
C* SIGE - THE COVARIANCE MATRIX
C* W(I) - THE VECTOR SCORE STATISTIC
      SUBROUTINE P4(K,NT1,YY,XX,SIGE,X1,Y1,Y2,W,ANS,V,S,TOT)
C***
C**   THIS SUBROUTINE COMPUTE THE STATISTIC S.
C**   A GENERALIZED KRUSKAL=WALLIS TEST FOR COMPARING
C**   K SAMPLES FROM P.583 OF BRESLOW.
C**
C**   SIGMA MATRIX HAS RANK K=1 PROVIDED THAT EACH OF
C**   THE K SAMPLES CONTAINS AT LEAST ONE UNCENSORED
C**   OBSERVATION. FOR SUCH SIGMA THERE EXIST K=1 VECTOR
C**   X(I)=(XX(I,1),...,XX(I,K))' (I=1,..., K=1) SUCH
C**   THAT X'(I)(SIGMA)X(J) EQUALS ONE OR ZERO ACCORDING
C**   AS I AND J ARE EQUAL OR UNEQUAL. CONSEQUENTLY THE
C**   STATISTICS
C**        S(I)=N**(=3/2) * (X'(I)W)
C**        WHERE W=(W(1),..., W(K))'
C**   WILL BE ASYMPTOTICALLY UNCORRELATED WITH MEAN 0
C**   AND UNIT VARIANCE. N**(=3/2)W HAS ASYMPTOTICALLY
C**   A MULTIDIMENSIONAL NORMAL DISTRIBUTION. FROM
C**   THIS IT FOLLOWS THAT
C**        S=SUM I=1, K=1 OF (S(I)**2)
C**   IS ASYMPTOTICALLY DISTRIBUTED IN A CHI=SQUARED
C**   DISTRIBUTION WITH K=1 DEGREES OF FREEDOM. THE
C**   STATISTICS S WILL BE USED TO TEST THE HYPOTHESIS.
C**
C**   IN PRACTICE, THE X(I) ARE FOUND BY MEANS OF THE
C**   GRAM=SCHMIDT ORTHOGONALIZATION PROCESS WITH THE
C**   INNER PRODUCT OF TWO VECTORS A AND B BY
C**   <A,B>=A'(SIGMA)B. THE K=1 VECTORS
C**   Y(1)=(1,0,...,0)', Y(2)=(1,1,0,...,0)',...,
C**   Y(K=1)=(1,...,1,0)'
C**   MAY BE USED AS A STARTING POINT FOR THE GRAM=
C**   SCHMIDT PROCESS.
C**
C**   GRAM=SCHMIDT PROCESS
C**   1.   X(1)=Y(1)/(//Y(1)//), //Y(1)//=SQRT(<Y(1),Y(1)>)
C**   2.   Z(2)=Y(2)=<Y(2),X(1)>X(1)
C**        X(2)=Z(2)/(//Z(2)//), //Z(2)//=SQRT(<Z(2),Z(2)>)
C**   3.   Z(3)=Y(3)=<Y(3),X(1)>X(1)=<Y(3),X(2)>X(2)
C**        X(3)=Z(3)/(//Z(3)//), //Z(3)//=SQRT(<Z(3),Z(3)>)
C**   .
C**   .
C**   .
C**   K=1.   Z(K=1)=Y(K=1)=<Y(K=1),X(1)>X(1)
C**                        =<Y(K=1),X(2)>X(2)
```

```
C**                     - ... -<Y(K-1),X(K-2)>X(K-2)
C**          X(K-1)=Z(K-1)/(//Z(K-1)//)
C***
      DIMENSION YY(10,10),XX(10,10),SIGE(10,10),X1(10),Y1(10),
     *          Y2(10),W(10),ANS(10),V(10),S(10)
      REAL NT1
      TOT=0
      AN=SQRT(NT1)
      WRITE(6,470)
  470 FORMAT(1X,///,' ----- GRAM-SCHMIDT ORTHOGONALIZATION',
     *        ' VECTORS -----')
      DO 410 I=1,K-1
         DO 420 J=1,K
         YY(I,J)=0
         IF(J.LE.I) YY(I,J)=1
         Y1(J)=YY(I,J)
  420    CONTINUE
         DO 425 N=1,I
            DO 426 J=1,K
            Y2(J)=XX(N,J)
  426       CONTINUE
         VA=VAL(K,Y1,Y2,ANS,SIGE)
         V(N)=VA
  425    CONTINUE
         DO 430 L=1,I
         DO 430 J=1,K
         Y1(J)=Y1(J)-V(L)*XX(L,J)
  430    CONTINUE
         DO 440 J=1,K
         Y2(J)=Y1(J)
         YY(I,J)=Y1(J)
  440    CONTINUE
      VA=VAL(K,Y1,Y2,ANS,SIGE)
      DEN=SQRT(VA)
      SUM=0
         DO 450 J=1,K
         XX(I+1,J)=YY(I,J)/DEN
         X1(J)=XX(I+1,J)
         SUM=SUM+X1(J)*W(J)
  450    CONTINUE
      S(I)=SUM/AN
      TOT=TOT+S(I)*S(I)
  410 CONTINUE
      WRITE(6,465) (I,I=1,K-1)
  465 FORMAT(1X,/,8(1X,I4,4X))
      DO 455 J=1,K
  455 WRITE(6,460) (XX(I+1,J),I=1,K-1)
  460 FORMAT(1X,/,8(1X,F8.5))
      RETURN
      END
C*
C*
      FUNCTION VAL(K,Y1,Y2,ANS,SIGE)
C***
C**   CALCULATE THE INNER PRODUCT OF TWO VECTORS A AND B
C**   BY < A, B> = A' (SIGMA) B
C***
      DIMENSION Y1(10),Y2(10),ANS(10),SIGE(10,10)
      DO 460 M=1,K-1
      ANS(M)=0
      DO 460 J=1,K
```

```
  460 ANS(M)=ANS(M)+Y1(J)*SIGE(J,M)
      VAL=0
      DO 470 I=1,K
  470 VAL=VAL+ANS(I)*Y2(I)
      RETURN
      END
C* NT1 = THE CUBIC OF TOTAL OBSERVATIONS
```

```
100   3
200   15
300   1,6162,1
400   3,3108,1
500   6,4986,1
600   8,7001,1
700   9,5813,1
800   10,3311,1
900   11,3729,1
1000  11,6698,1
1100  14,7173,1
1200  20,0
1300  20,0
1400  20,0
1500  20,0
1600  20,0
1700  20,0
1800  12
1900  0,3113,1
2000  0,3193,1
2100  0,5793,1
2200  0,9393,1
2300  1,4585,1
2400  2,6174,1
2500  2,8732,1
2600  6,5377,1
2700  12,9082,1
2800  20,0
2900  20,0
3000  20,0
3100  18
3200  1,0977,1
3300  1,5408,1
3400  1,7251,1
3500  2,1613,1
3600  2,8869,1
3700  3,0430,1
3800  3,1428,1
3900  3,5553,1
4000  3,9252,1
4100  4,2688,1
4200  4,3705,1
4300  6,9837,1
4400  7,7971,1
4500  7,9731,1
4600  8,2069,1
4700  9,2274,1
4800  11,2849,1
4900  20,0
5000  3,0,05
5100  1,2
5200  1,3
5300  2,3
```

```
****************************************************

 NT=THE TOTAL SAMPLE SIZE=  45
NT1=THE CUBIC OF TOTAL SAMPLE SIZE=      91125

===== VECTOR SCORE STATISTIC =====

W( 1)=   258,

W( 2)=  -119,

W( 3)=  -139,

THE SUM OF W(I) EQUAL 0,

===== SIGMA MATRIX =====

 0,0761 -0,0272 -0,0489

-0,0272  0,0504 -0,0232

-0,0489 -0,0232  0,0721

===== GRAM-SCHMIDT ORTHOGONALIZATION VECTORS =====

   1         2

 3,62437  1,77370

 0,00000  4,95737

 0,00000  0,00000

===== TESTING HYPOTHESIS =====

S**=    9,912700

  P[CHI-SQUARE( 2) >=    9,912700 ] =0,007039

 S*=  10,352355

  P[CHI-SQUARE( 2) >=  10,352355 ] =0,005650

 S=    9,787615
```

```
  P[CHI-SQUARE( 2) >=    9,787615 ] =0,007493

S** AS A LOWER BOUND TO S*, S AND S* WILL BE
ASYMPTOTICALLY EQUIVALENT STATISTICS,
S* IS COMPUTATIONALLY SIMPLER THAN S,
HOWEVER, ONLY S WILL BE AN ASYMPTOTICALLY
VALID STATISTIC UNDER HYPOTHESIS,


*****************************************************


******************************************************


  ----- 1 VS 2 -----


 NT=THE TOTAL SAMPLE SIZE=  27
NT1=THE CUBIC OF TOTAL SAMPLE SIZE=      19683


----- VECTOR SCORE STATISTIC -----

W( 1)=    82,

W( 2)=   -82,

THE SUM OF W(I) EQUAL 0,


----- SIGMA MATRIX -----

 0,0670 -0,0670

-0,0670  0,0670


----- GRAM-SCHMIDT ORTHOGONALIZATION VECTORS -----

   1

 3,86299

 0,00000


----- TESTING HYPOTHESIS -----

S**=   4,150617

  P[CHI-SQUARE(1) >= 4,150617 ] =0,041619
```

```
S*=   4,540970

 P(CHI-SQUARE(1) >= 4,540970 ) =0,033093


S=   5,097801

 P(CHI-SQUARE(1) >= 5,097801 ) =0,023956
USING BONFERRONI INEQUALITY
REJECT 0,023956 IF 0,023956 < 0,05/3=0,016667


S** AS A LOWER BOUND TO S*, S AND S* WILL BE
ASYMPTOTICALLY EQUIVALENT STATISTICS,
S* IS COMPUTATIONALLY SIMPLER THAN S,
HOWEVER, ONLY S WILL BE AN ASYMPTOTICALLY
VALID STATISTIC UNDER HYPOTHESIS,


BONFERRONI CRITICAL VALUE
(ASSUMES ALPHA= 0,05, 3 PAIRWISE COMPARISONS)
=CHI-SQUARE FOR 1DF, (1 - 0,05/3)100(TH)%

 P(CHI-SQUARE(1) <= 5,737029 ) =0,983333


**************************************************


***************************************************


 ===== 1 VS 3 =====


 NT=THE TOTAL SAMPLE SIZE=  33
NT1=THE CUBIC OF TOTAL SAMPLE SIZE=     35937


===== VECTOR SCORE STATISTIC =====

W( 1)=   176,

W( 2)=  -176,

THE SUM OF W(I) EQUAL 0,


===== SIGMA MATRIX =====

 0,0734 -0,0734

-0,0734  0,0734
```

```
----- GRAM-SCHMIDT ORTHOGONALIZATION VECTORS -----

   1

 3,69161

 0,00000


----- TESTING HYPOTHESIS -----

S**=  10,429630

  P[CHI-SQUARE(1) >=10,429630 ] =0,001240


 S*=  11,008594

  P[CHI-SQUARE(1) >=11,008594 ] =0,000907


 S=  11,746682

  P[CHI-SQUARE(1) >=11,746682 ] =0,000610
 USING BONFERRONI INEQUALITY
REJECT 0,000610 IF 0,000610 < 0,05/3=0,016667


S** AS A LOWER BOUND TO S*, S AND S* WILL BE
ASYMPTOTICALLY EQUIVALENT STATISTICS,
S* IS COMPUTATIONALLY SIMPLER THAN S,
HOWEVER, ONLY S WILL BE AN ASYMPTOTICALLY
VALID STATISTIC UNDER HYPOTHESIS,


BONFERRONI CRITICAL VALUE
(ASSUMES ALPHA= 0,05, 3 PAIRWISE COMPARISONS)
=CHI-SQUARE FOR 1DF, (1 - 0,05/3)100(TH)%

  P[CHI-SQUARE(1) <= 5,737029 ] =0,983333


*****************************************************


*****************************************************


 ----- 2 VS 3 -----


 NT=THE TOTAL SAMPLE SIZE=  30
NT1=THE CUBIC OF TOTAL SAMPLE SIZE=      27000
```

```
===== VECTOR SCORE STATISTIC =====

W( 1)=    -37.

W( 2)=     37.

THE SUM OF W(I) EQUAL 0.


===== SIGMA MATRIX =====

 0.0694 -0.0694

-0.0694  0.0694


===== GRAM-SCHMIDT ORTHOGONALIZATION VECTORS =====

    1

 3.79676

 0.00000


===== TESTING HYPOTHESIS =====

S**=    0.633796

  P[CHI-SQUARE(1) >= 0.633796 ] =0.425966

 S*=    0.667857

  P[CHI-SQUARE(1) >= 0.667857 ] =0.413800

 S=    0.730913

  P[CHI-SQUARE(1) >= 0.730913 ] =0.392587
 USING BONFERRONI INEQUALITY
REJECT 0.392587 IF 0.392587 < 0.05/3=0.016667

S** AS A LOWER BOUND TO S*, S AND S* WILL BE
ASYMPTOTICALLY EQUIVALENT STATISTICS.
S* IS COMPUTATIONALLY SIMPLER THAN S.
HOWEVER, ONLY S WILL BE AN ASYMPTOTICALLY
VALID STATISTIC UNDER HYPOTHESIS.


BONFERRONI CRITICAL VALUE
(ASSUMES ALPHA= 0.05, 3 PAIRWISE COMPARISONS)
=CHI-SQUARE FOR 1DF, (1 - 0.05/3)100(TH)%

  P[CHI-SQUARE(1) <= 5.737029 ] =0.983333
```

****

```
100   2
200   25
300   3,1,4,1,4,1,4,1,4,1,4,1,4,1,4,1,4,1,4,1,4,1,4,1,4,1,4,1,
400   4,1,4,1,4,1,4,1,4,1,4,1,4,1,4,1,4,1,5,1
500   200
600   9,1,9,1,9,1,9,1,9,1,9,1,9,1,9,1,9,1,10,1,10,1,12,1,13,1,14,1,
700   15,1,15,1,20,1,31,1,32,1,33,1,50,1,50,1,55,1,56,1,62,1,
800   9,1,9,1,9,1,9,1,10,1,10,1,10,1,10,1,10,1,10,1,11,1,11,1,14,1,
900   15,1,16,1,18,1,22,1,23,1,23,1,24,1,30,1,37,1,47,1,100,0,100,0,
1000  8,1,8,1,11,1,11,1,12,1,13,1,13,1,13,1,14,1,16,1,17,1,17,1,
1100  18,1,18,1,18,1,20,1,21,1,23,1,25,1,25,1,25,1,27,1,28,1,
1200  29,1,43,1,
1300  18,1,19,1,19,1,22,1,24,1,24,1,24,1,26,1,29,1,29,1,30,1,30,1,
1400  30,1,33,1,100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,
1500  100,0,100,0,100,0,
1600  9,1,21,1,25,1,29,1,40,1,45,1,46,1,52,1,80,1,86,1,86,1,87,1,
1700  87,1,100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,
1800  100,0,100,0,100,0,
1900  24,1,100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,
2000  100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,
2100  100,0,100,0,100,0,100,0,100,0,
2200  12,1,18,1,18,1,20,1,27,1,36,1,47,1,50,1,52,1,58,1,68,1,100,0,
2300  100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,
2400  100,0,100,0,100,0,
2500  21,1,100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,
2600  100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,
2700  100,0,100,0,100,0,100,0,100,0,
```

```
*****************************************************

 NT=THE TOTAL SAMPLE SIZE= 225
NT1=THE CUBIC OF TOTAL SAMPLE SIZE=  11390625

===== VECTOR SCORE STATISTIC =====

W( 1)= -5000.

W( 2)=  5000.

THE SUM OF W(I) EQUAL 0.

===== SIGMA MATRIX =====

 0.0008 -0.0008

-0.0008  0.0008

===== GRAM-SCHMIDT ORTHOGONALIZATION VECTORS =====

    1

34.81047

 0.00000

===== TESTING HYPOTHESIS =====

S**=  66.666667

  P[CHI-SQUARE( 1) >=  66.666667 ] =0.000000

 S*=  74.892909

  P[CHI-SQUARE( 1) >=  74.892909 ] =0.000000

 S=2659.574468

  P[CHI-SQUARE( 1) >=2659.574468 ] =0.000000

S** AS A LOWER BOUND TO S*, S AND S* WILL BE
ASYMPTOTICALLY EQUIVALENT STATISTICS.
S* IS COMPUTATIONALLY SIMPLER THAN S.
```

HOWEVER, ONLY S WILL BE AN ASYMPTOTICALLY
VALID STATISTIC UNDER HYPOTHESIS.

****

```
100     2
200     25
300     9,1,9,1,9,1,9,1,9,1,9,1,9,1,9,1,9,1,10,1,10,1,12,1,13,1,14,1,
400     15,1,15,1,20,1,31,1,32,1,33,1,50,1,50,1,55,1,56,1,62,1,
500     175
600     9,1,9,1,9,1,9,1,10,1,10,1,10,1,10,1,10,1,10,1,11,1,11,1,14,1,
700     15,1,16,1,18,1,22,1,23,1,23,1,24,1,30,1,37,1,47,1,100,0,100,0,
800     8,1,8,1,11,1,11,1,12,1,13,1,13,1,13,1,14,1,16,1,17,1,17,1,
900     18,1,18,1,18,1,20,1,21,1,23,1,25,1,25,1,25,1,27,1,28,1,
1000    29,1,43,1,
1100    18,1,19,1,19,1,22,1,24,1,24,1,24,1,26,1,29,1,29,1,30,1,30,1,
1200    30,1,33,1,100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,
1300    100,0,100,0,100,0,
1400    9,1,21,1,25,1,29,1,40,1,45,1,46,1,52,1,80,1,86,1,86,1,87,1,
1500    87,1,100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,
1600    100,0,100,0,100,0,
1700    24,1,100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,
1800    100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,
1900    100,0,100,0,100,0,100,0,100,0,
2000    12,1,18,1,18,1,20,1,27,1,36,1,47,1,50,1,52,1,58,1,68,1,100,0,
2100    100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,
2200    100,0,100,0,100,0,
2300    21,1,100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,
2400    100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,
2500    100,0,100,0,100,0,100,0,100,0
```

```
*****************************************************

 NT=THE TOTAL SAMPLE SIZE= 200
NT1=THE CUBIC OF TOTAL SAMPLE SIZE=   8000000

----- VECTOR SCORE STATISTIC -----

W( 1)= -2738.

W( 2)=  2738.

THE SUM OF W(I) EQUAL 0.

----- SIGMA MATRIX -----

 0.0189 -0.0189

-0.0189  0.0189

----- GRAM-SCHMIDT ORTHOGONALIZATION VECTORS -----

    1

 7.27181

 0.00000

----- TESTING HYPOTHESIS -----

S**=  25.702779

  P[CHI-SQUARE( 1) >=  25.702779 ] =0.000000

 S*=  29.037435

  P[CHI-SQUARE( 1) >=  29.037435 ] =0.000000

 S=  49.552139

  P[CHI-SQUARE( 1) >=  49.552139 ] =0.000000

S** AS A LOWER BOUND TO S*. S AND S* WILL BE
ASYMPTOTICALLY EQUIVALENT STATISTICS.
S* IS COMPUTATIONALLY SIMPLER THAN S.
```

HOWEVER, ONLY S WILL BE AN ASYMPTOTICALLY
VALID STATISTIC UNDER HYPOTHESIS.

****

```
 100    9
 200    25
 300    3,1,4,1,4,1,4,1,4,1,4,1,4,1,4,1,4,1,4,1,4,1,4,1,4,1,4,1,4,1,
 400    4,1,4,1,4,1,4,1,4,1,4,1,4,1,4,1,4,1,5,1
 500    25
 600    9,1,9,1,9,1,9,1,9,1,9,1,9,1,9,1,9,1,10,1,10,1,12,1,13,1,14,1,
 700    15,1,15,1,20,1,31,1,32,1,33,1,50,1,50,1,55,1,56,1,62,1
 800    25
 900    9,1,9,1,9,1,9,1,10,1,10,1,10,1,10,1,10,1,10,1,11,1,11,1,14,1,
1000    15,1,16,1,18,1,22,1,23,1,23,1,24,1,30,1,37,1,47,1,100,0,100,0
1100    25
1200    8,1,8,1,11,1,11,1,12,1,13,1,13,1,13,1,14,1,16,1,17,1,17,1
1300    18,1,18,1,18,1,20,1,21,1,23,1,25,1,25,1,25,1,27,1,28,1,
1400    29,1,43,1
1500    25
1600    18,1,19,1,19,1,22,1,24,1,24,1,24,1,26,1,29,1,29,1,30,1,30,1,
1700    30,1,33,1,100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,
1800    100,0,100,0,100,0
1900    25
2000    9,1,21,1,25,1,29,1,40,1,45,1,46,1,52,1,80,1,86,1,86,1,87,1,
2100    87,1,100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,
2200    100,0,100,0,100,0
2300    25
2400    24,1,100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,
2500    100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,
2600    100,0,100,0,100,0,100,0,100,0
2700    25
2800    12,1,18,1,18,1,20,1,27,1,36,1,47,1,50,1,52,1,58,1,68,1,100,0,
2900    100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,
3000    100,0,100,0,100,0
3100    25
3200    21,1,100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,
3300    100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,100,0,
3400    100,0,100,0,100,0,100,0,100,0
3500    6,0,05
3600    3,4
3700    3,8
3800    4,6
3900    5,7
4000    7,8
4100    8,9
```

```
**************************************************

 NT=THE TOTAL SAMPLE SIZE= 225
NT1=THE CUBIC OF TOTAL SAMPLE SIZE=  11390625

----- VECTOR SCORE STATISTIC -----
W( 1)= -5000.
W( 2)= -2113.
W( 3)= -1944.
W( 4)= -1945.
W( 5)=  1055.
W( 6)=  1694.
W( 7)=  3272.
W( 8)=  1727.
W( 9)=  3254.
THE SUM OF W(I) EQUAL 0.

----- SIGMA MATRIX -----
 0.0008 -0.0001 -0.0001 -0.0001 -0.0001 -0.0001 -0.0001 -0.0001 -0.0001
-0.0001  0.0230 -0.0025 -0.0027 -0.0034 -0.0035 -0.0036 -0.0035 -0.0036
-0.0001 -0.0025  0.0234 -0.0029 -0.0035 -0.0035 -0.0037 -0.0035 -0.0037
-0.0001 -0.0027 -0.0029  0.0251 -0.0038 -0.0038 -0.0039 -0.0038 -0.0039
-0.0001 -0.0034 -0.0035 -0.0038  0.0347 -0.0058 -0.0062 -0.0057 -0.0062
-0.0001 -0.0035 -0.0035 -0.0038 -0.0058  0.0364 -0.0068 -0.0062 -0.0067
-0.0001 -0.0036 -0.0037 -0.0039 -0.0062 -0.0068  0.0382 -0.0066 -0.0073
-0.0001 -0.0035 -0.0035 -0.0038 -0.0057 -0.0062 -0.0066  0.0360 -0.0066
-0.0001 -0.0036 -0.0037 -0.0039 -0.0062 -0.0067 -0.0073 -0.0066  0.0382

----- GRAM-SCHMIDT ORTHOGONALIZATION VECTORS -----
```

1	2	3	4	5	6	7	8
34,81047	0,82469	0,91310	1,01961	1,09893	1,37721	1,94955	3,29697
0,00000	6,59750	0,72431	0,85743	1,07114	1,36545	1,96068	3,27845
0,00000	0,00000	6,58051	0,88252	1,08137	1,36521	1,95909	3,28122
0,00000	0,00000	0,00000	6,41692	1,09559	1,36753	1,95578	3,28701
0,00000	0,00000	0,00000	0,00000	5,54332	1,35728	1,97038	3,26213
0,00000	0,00000	0,00000	0,00000	0,00000	5,56225	1,97524	3,25345
0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	5,77519	3,24642
0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	6,76711
0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000

===== TESTING HYPOTHESIS =====

S**= 154,753470

P(CHI-SQUARE(8) >= 154,753470) =0,000000

S*= 173,849062

P(CHI-SQUARE(8) >= 173,849062) =0,000000

S=2766,360697

P(CHI-SQUARE(8) >=2766,360697) =0,000000

S** AS A LOWER BOUND TO S*, S AND S* WILL BE
ASYMPTOTICALLY EQUIVALENT STATISTICS,
S* IS COMPUTATIONALLY SIMPLER THAN S,
HOWEVER, ONLY S WILL BE AN ASYMPTOTICALLY
VALID STATISTIC UNDER HYPOTHESIS,

****

****

===== 3 VS 4 =====

NT=THE TOTAL SAMPLE SIZE= 50
NT1=THE CUBIC OF TOTAL SAMPLE SIZE= 125000

```
===== VECTOR SCORE STATISTIC =====

W( 1)=  -106.

W( 2)=   106.

THE SUM OF W(I) EQUAL 0.


===== SIGMA MATRIX =====

 0.0730 -0.0730

-0.0730  0.0730


===== GRAM-SCHMIDT ORTHOGONALIZATION VECTORS =====

    1

 3.70177

 0.00000


===== TESTING HYPOTHESIS =====

S**=    1.078656

  P[CHI-SQUARE(1) >= 1.078656 ] =0.298998

 S*=    1.200299

  P[CHI-SQUARE(1) >= 1.200299 ] =0.273262

 S=    1.231747

  P[CHI-SQUARE(1) >= 1.231747 ] =0.267067
 USING BONFERRONI INEQUALITY
REJECT 0.267067 IF 0.267067 < 0.05/6=0.008333

S** AS A LOWER BOUND TO S*, S AND S* WILL BE
ASYMPTOTICALLY EQUIVALENT STATISTICS.
S* IS COMPUTATIONALLY SIMPLER THAN S.
HOWEVER, ONLY S WILL BE AN ASYMPTOTICALLY
VALID STATISTIC UNDER HYPOTHESIS.


BONFERRONI CRITICAL VALUE
(ASSUMES ALPHA= 0.05, 6 PAIRWISE COMPARISONS)
=CHI-SQUARE FOR 1DF, (1 - 0.05/6)100(TH)%

  P[CHI-SQUARE(1) <= 6.981694 ] =0.991667
```

```
****************************************************

****************************************************

 ===== 3 VS 8 =====

 NT=THE TOTAL SAMPLE SIZE=  50
NT1=THE CUBIC OF TOTAL SAMPLE SIZE=     125000

===== VECTOR SCORE STATISTIC =====

W( 1)=  -476.

W( 2)=   476.

THE SUM OF W(I) EQUAL 0.

===== SIGMA MATRIX =====

 0.0634 -0.0634

-0.0634  0.0634

===== GRAM-SCHMIDT ORTHOGONALIZATION VECTORS =====

   1

 3.97251

 0.00000

===== TESTING HYPOTHESIS =====

S**=  21.751296

  P[CHI-SQUARE(1) >=21.751296 ] =0.000003

 S*=  24.494703

  P[CHI-SQUARE(1) >=24.494703 ] =0.000001

 S=  28.604469
```

```
  P[CHI-SQUARE(1) >=28.604469 ] =0.000000
 USING BONFERRONI INEQUALITY
REJECT 0.000000 IF 0.000000 < 0.05/6=0.008333

S** AS A LOWER BOUND TO S*, S AND S* WILL BE
ASYMPTOTICALLY EQUIVALENT STATISTICS.
S* IS COMPUTATIONALLY SIMPLER THAN S.
HOWEVER, ONLY S WILL BE AN ASYMPTOTICALLY
VALID STATISTIC UNDER HYPOTHESIS.


BONFERRONI CRITICAL VALUE
(ASSUMES ALPHA= 0.05, 6 PAIRWISE COMPARISONS)
=CHI-SQUARE FOR 1DF, (1 - 0.05/6)100(TH)%

  P[CHI-SQUARE(1) <= 6.981694 ] =0.991667


*****************************************************


******************************************************

  ----- 4 VS 6 -----


 NT=THE TOTAL SAMPLE SIZE=  50
NT1=THE CUBIC OF TOTAL SAMPLE SIZE=     125000


----- VECTOR SCORE STATISTIC -----

W( 1)=  -546.

W( 2)=   546.

THE SUM OF W(I) EQUAL 0.


----- SIGMA MATRIX -----

 0.0622 -0.0622

-0.0622  0.0622


----- GRAM-SCHMIDT ORTHOGONALIZATION VECTORS -----

    1

 4.01067
```

```
 0,00000

===== TESTING HYPOTHESIS =====

S**=  28,619136

  P[CHI-SQUARE(1) >=28,619136 ] =0,000000

 S*=  30,806655

  P[CHI-SQUARE(1) >=30,806655 ] =0,000000

 S=  38,362630

  P[CHI-SQUARE(1) >=38,362630 ] =0,000000
 USING BONFERRONI INEQUALITY
REJECT 0,000000 IF 0,000000 < 0,05/6=0,008333

S** AS A LOWER BOUND TO S*, S AND S* WILL BE
ASYMPTOTICALLY EQUIVALENT STATISTICS,
S* IS COMPUTATIONALLY SIMPLER THAN S,
HOWEVER, ONLY S WILL BE AN ASYMPTOTICALLY
VALID STATISTIC UNDER HYPOTHESIS,


BONFERRONI CRITICAL VALUE
(ASSUMES ALPHA= 0,05, 6 PAIRWISE COMPARISONS)
=CHI-SQUARE FOR 1DF, (1 - 0,05/6)100(TH)%

  P[CHI-SQUARE(1) <= 6,981694 ] =0,991667


**************************************************


***************************************************

  ===== 5 VS 7 =====


 NT=THE TOTAL SAMPLE SIZE=  50
NT1=THE CUBIC OF TOTAL SAMPLE SIZE=     125000


===== VECTOR SCORE STATISTIC =====

W( 1)=  -322,

W( 2)=   322,
```

```
THE SUM OF W(I) EQUAL 0.

===== SIGMA MATRIX =====

 0.0497 =0.0497
=0.0497  0.0497

===== GRAM-SCHMIDT ORTHOGONALIZATION VECTORS =====

   1
 4.48688
 0.00000

===== TESTING HYPOTHESIS =====

S**=   9.953664
  P(CHI-SQUARE(1) >= 9.953664 ) =0.001605

 S*=  16.058234
  P(CHI-SQUARE(1) >=16.058234 ) =0.000061

 S=  16.698985
  P(CHI-SQUARE(1) >=16.698985 ) =0.000044
 USING BONFERRONI INEQUALITY
REJECT 0.000044 IF 0.000044 < 0.05/6=0.008333

S** AS A LOWER BOUND TO S*, S AND S* WILL BE
ASYMPTOTICALLY EQUIVALENT STATISTICS.
S* IS COMPUTATIONALLY SIMPLER THAN S.
HOWEVER, ONLY S WILL BE AN ASYMPTOTICALLY
VALID STATISTIC UNDER HYPOTHESIS.

BONFERRONI CRITICAL VALUE
(ASSUMES ALPHA= 0.05, 6 PAIRWISE COMPARISONS)
=CHI-SQUARE FOR 1DF, (1 - 0.05/6)100(TH)%

  P(CHI-SQUARE(1) <= 6.981694 ) =0.991667

**************************************************
```

```
*****************************************************

  ===== 7 VS 8 =====


 NT=THE TOTAL SAMPLE SIZE=  50
NT1=THE CUBIC OF TOTAL SAMPLE SIZE=     125000


===== VECTOR SCORE STATISTIC =====

W( 1)=   247.

W( 2)=  -247.

THE SUM OF W(I) EQUAL 0.


===== SIGMA MATRIX =====

 0.0447 -0.0447

-0.0447  0.0447


===== GRAM-SCHMIDT ORTHOGONALIZATION VECTORS =====

   1

 4.73090

 0.00000


===== TESTING HYPOTHESIS =====

S**=   5.856864

  P[CHI-SQUARE(1) >= 5.856864 ] =0.015516

 S*=  10.724500

  P[CHI-SQUARE(1) >=10.724500 ] =0.001057

 S=  10.923724

  P[CHI-SQUARE(1) >=10.923724 ] =0.000949
 USING BONFERRONI INEQUALITY
REJECT 0.000949 IF 0.000949 < 0.05/6=0.008333

S** AS A LOWER BOUND TO S*, S AND S* WILL BE
ASYMPTOTICALLY EQUIVALENT STATISTICS.
```

```
S* IS COMPUTATIONALLY SIMPLER THAN S.
HOWEVER, ONLY S WILL BE AN ASYMPTOTICALLY
VALID STATISTIC UNDER HYPOTHESIS.


BONFERRONI CRITICAL VALUE
(ASSUMES ALPHA= 0.05, 6 PAIRWISE COMPARISONS)
=CHI-SQUARE FOR 1DF, (1 - 0.05/6)100(TH)%

  P[CHI-SQUARE(1) <= 6.981694 ] =0.991667


***************************************************


****************************************************

  ----- 8 VS 9 -----


 NT=THE TOTAL SAMPLE SIZE=  50
NT1=THE CUBIC OF TOTAL SAMPLE SIZE=     125000


----- VECTOR SCORE STATISTIC -----

W( 1)=  -247.

W( 2)=   247.

THE SUM OF W(I) EQUAL 0.


----- SIGMA MATRIX -----

 0.0447 -0.0447

-0.0447  0.0447


----- GRAM-SCHMIDT ORTHOGONALIZATION VECTORS -----

   1

 4.73090

 0.00000


----- TESTING HYPOTHESIS -----

S**=   5.856864
```

```
 P[CHI-SQUARE(1) >= 5,856864 ] =0,015516

S*=  10,724500

 P[CHI-SQUARE(1) >=10,724500 ] =0,001057

S=  10,923724

 P[CHI-SQUARE(1) >=10,923724 ] =0,000949
 USING BONFERRONI INEQUALITY
REJECT 0,000949 IF 0,000949 < 0,05/6=0,008333

S** AS A LOWER BOUND TO S*, S AND S* WILL BE
ASYMPTOTICALLY EQUIVALENT STATISTICS,
S* IS COMPUTATIONALLY SIMPLER THAN S,
HOWEVER, ONLY S WILL BE AN ASYMPTOTICALLY
VALID STATISTIC UNDER HYPOTHESIS,


BONFERRONI CRITICAL VALUE
(ASSUMES ALPHA= 0,05, 6 PAIRWISE COMPARISONS)
=CHI-SQUARE FOR 1DF, (1 - 0,05/6)100(TH)%

 P[CHI-SQUARE(1) <= 6,981694 ] =0,991667


****************************************************
```